EUKARYOTIC CELL EXPRESSING PHYCOCYANIN

[0001] The invention relates to recombinantly prepared phycocyanin and functionally active parts thereof and claims the priority of European patent application 04 001 504.2, the contents of which are incorporated by reference.

[0002] In all areas of modern biology - e.g. in DNA replication, transcription, translation or control of the cell cycle, and signal transductions, the interaction between proteins plays an essential role. Consequently, elucidation of these processes is no longer the domain solely of biochemists. On the contrary, it is now also necessary for geneticists and molecular biologists or biophysicists to be concerned with the elucidation of protein-protein interactions.

[0003] The interactions between proteins are conventionally identified or quantified with the aid of biochemical methods such as affinity chromatography or immunoprecipitation. These methods are frequently combined with genetic or molecular biology methods. If, for example, a protein which is sought is expressed by a bacteriophage/cDNA library, the interaction can in some cases be detected using a specific protein probe directly in the lysate of a positive plaque.

[0004] Many of the methods for investigating protein-protein interactions are, however, based on the detection of these interactions *in vitro*. Thus, on the one hand, the applicability of the results generated in this way to *in vivo* conditions is doubtful. Furthermore, it is possible to detect only interactions of binding partners whose dissociation constants are not too large. Otherwise, the interactions to be detected would not withstand the usually stringent and time-consuming washing protocols. However, these are normally necessary in order to be able to reduce interfering background signals to an acceptable extent. The development of systems for investigating protein-protein interactions *in vivo* therefore represents a special challenge.

[0005] In the recent past, many scientists have used the so-called green fluorescent protein (GFP) from *Aequorea victoria* for studying cell biology. GFP is a fluorescent protein which has now been successfully cloned and expressed in a large number of cells and organisms - e.g. in bacteria, plant cells and animal cells. GFP can thus be used in cell biology for example for localizing proteins, for detecting successful transfection or as reporter gene for gene expression.

[0006] GFP shows an absorption spectrum with two peaks, namely a first peak at 395 nm and a second at 470 nm. Tyr^{66} is responsible therefor, being converted in its hydroxy form into the phenolate form by excitation at 395 nm. The latter emits, after appropriate excitation, light with a wavelength of 509 nm.

[0007] The emission and absorption spectrum of GFP is thus located exactly in the spectral ranges in which a large number of biological substances exhibit fluorescence. This natural fluorescence of biological systems - also known as autofluorescence - is derived to not inconsiderable extent from flavin derivatives or from NADH and NADPH derivatives. Thus, for example, flavins are excited by light of 380-490 nm and emit light of a wavelength of 520-560 nm. NADH and NADPH show absorption maxima at 360-390 nm, while they exhibit the highest emission at 440-470 nm.

[0008] This autofluorescence therefore overlaps with the spectrum typical of the GFP protein, and is responsible for a high background noise when GFP is used in *in vivo* studies. This background noise not infrequently leads to a loss of clarity and informativeness of the experimental results.

[0009] Methods have been developed in the past for distinguishing autofluorescence from GFP fluorescence. These methods are based for example on optimization of the optical filters used for the fluorescence detection or on sequential excitation of the sample first with a wavelength optimized for excitation of GFP, and subsequently with a wavelength optimized for autofluorescence. The

difference between the resulting measured signals is regarded as fluorescence of GFP alone. This latter method is known as dual wavelength correction. Other methods, for example microscopy image correction, are also based ultimately on a correction of the GFP fluorescence signal for the level of autofluorescence.

[0010] Besides these, elaborate optimizations of the actual evaluation unit have been undertaken. Thus, for example, confocal laser microscopy or two-photon laser scan microscopy is employed to obtain a high resolution of the fluorescence signals and to be able to assign them thus as far as possible to individual specific three-dimensional or spatial structures within the biological sample. This last approach derives from observations that autofluorescence is usually not homogeneously distributed within the sample.

[0011] A majority of the considerations for improving the evaluation of a GFP signal in an *in vivo* sample is therefore aimed at the design of the measurement method and at optimization of the detection or evaluation unit. (Concerning this in detail see: Billinton N and A.W. Knight: Seeing the Wood through the Trees: A Review of Techniques for distinguishing GFP from Cellular endogenous Autofluorescence. Analytical Biochemistry 2001 (291), 175-197).

[0012] Besides these, only a few approaches to modifying the GFP by mutagenesis so that the absorption and the emission is shifted into longer-wavelength light, and thus leaves the critical region of autofluorescence, exist. Thus, for example, the so-called red fluorescent protein (RFP or DsRed) from *Discosoma* with an emission maximum at 583 nm has been isolated, and the further development thereof, mRFP1, shows an absorption maximum at 619 nm. It is true that this variant does not, in contrast to the original RFP - have an interfering second peak at 490 nm. However, it is characterized by a lower photostability and signal strength by comparison therewith (Zhang J., E. Campbell, A. Ting and Y. Tsien: Creating New Fluorescent Probes for Cell Biology, Molecular Cell Biology 2002 (3), pp. 906-918).

[0013] It is therefore an object of the present invention to provide an alternative fluorescent protein in particular for cell biology investigations, and a method for preparing it. This protein was intended in particular to exhibit an absorption maximum in the longer-wavelength range.

[0014] This object is achieved by the provision of a phycocyanin or functionally active parts thereof, where the phycocyanin and its parts according to the invention exhibit a glycosylation by comparison with native phycocyanin. Also provided is a transformed eukaryotic cell, in particular a yeast cell, which expresses the phycocyanin of the invention or its functionally active parts.

[0015] Phycocyanins are phycobiliproteins which are known as accessory photosynthesis pigments from cyanobacteria, red algae or cryptomonads. Some phycobiliproteins are present in cyanobacteria and red algae predominantly aggregated to large subunits, called phycobilisomes.

[0016] Phycobiliproteins belong to the family of homologous proteins which in each case consist of an αβ heterodimer closely linked together. Both subunits - both the α subunit and the β subunit may comprise one or more chromophores. The nature and the number of the chromophores covalently bonded to the protein unit (apo-protein) serves to classify the phycobiliproteins.

[0017] Phycocyanins have a single bilin unit on the α subunit and two bilins on the β subunit. By comparison therewith, for example, phycoerythrins are characterized by two or three bilins on the α subunit and three on the β subunit. Allophycocyanins by contrast in each case have only one bilin on each protein subunit.

[0018] As already mentioned, the bilins are differentiated on the basis of their structure. However, all bilins are consistently chromophores composed of a linear tetrapyrrole which is covalently bonded via a thioether linkage between the

C3' atom of the bilin and a cysteine residue of the protein unit to the apo-protein. The number of thioether linkages may, however, vary. Thus, some bilins have two thioether linkages to the apo-protein. It is likewise possible for the configuration of the C3' atom to be R or S.

[0019] The spectroscopic properties of the phycobiliproteins are substantially determined by the number and the position of the double bonds in the bilin. Thus, a distinction is made between phycocyanobilin (PCB), phycobiliviolin (PXB), phycoerythrobilin (PEB) and phycourobilin (PUB). The PEB and PUB having two thioether linkages to the apo-protein are called DiCysPEB and DiCysPUB, respectively.

[0020] The conformation of the bilin portion is therefore of great importance for the spectral properties of the phycobiliprotein. Thus, X-ray structure analyses show that the bilin in the phycobiliprotein usually assumes a maximally extended conformation. On the other hand, there is a relatively strong shift of the absorption maximum into the UV range when the bilin is present free in solution - whereby it may assume a conformation differing therefrom.

[0021] Phycocyanin is - as already mentioned - characterized by a bilin on the α subunit and two bilins on the β subunit. The bilin of the α subunit is always a PCB. The β subunit may have either two PCB or else one PCB and PEB. The former case comprises C-phycocyanin (C-PC), while the second molecule is called R-phycocyanin (R-PC). However, there is no substantial difference between the two PC variants in their spectral properties.

[0022] Most phycobiliproteins absorb in the 495-650 nm range. Phycocyanin shows an absorption maximum at 600-650 nm - depending on the literature, experimental conditions and initial organism. According to Tooley *et al* (Tooley A.J., Yuping A.C., Glazer A.N.: Biosynthesis of a fluorescent cyanobacterial C-Phycocyanin holo-α subunit in a heterologous host. PNAS 2001

vol. 19) 10560-10565), the Holo-CpcA (i.e. the α subunit of PC) shows an absorption maximum of 619 nm and a maximum emission at 641 nm.

[0023] These spectral properties of phycocyanin permit a clear discrimination of the actual fluorescence measured signal from the interfering background noise of autofluorescence and thus form the basis for its particular suitability as fluorescent marker for *in vivo* investigations. The autofluorescence derives - as already mentioned above - substantially from flavin or NADH or NADPH derivatives showing absorption and emission maxima at 360-560 nm.

[0024] The fundamental investigations for the biosynthesis of phycocyanin were done with the red alga *Cyanidium caldarium*. It became clear from this, through ^{14}C-labelings of the heme, that the initial precondition for biosynthesis of the phycocyanin is degradation of endogenous heme to biliverdin IXα by heme oxygenase 1 (HO1). This entails linearization of the tetrapyrrole ring. In further steps, the biliverdin is converted into 14,16-dihydrobiliverdin IXα and subsequently into 3Z-phycocyanobilin.

[0025] The precondition for these last two conversions is in each case the presence of reduced ferredoxin, and they are catalyzed by ferredoxin oxidoreductase. The PCB resulting therefrom is linked by the heterodimeric phycocyanin α PCB lyase to the α subunit of the phycobiliprotein (apo-CpcA). This results in the complete holo α subunit (holo-CpcA).

[0026] Thus, biosynthesis of holo-CpcA comprises initial degradation and conversion of the heme to PCB and subsequent enzymatic linkage thereof to the protein subunit. Biosynthesis of the holo β subunit (holo CpcB) proceeds correspondingly, with the linking of the bilin units to the CpcB presumably taking place non-enzymatically, at least in part. If both the holo-CpcA and the holo-CpcB are present, they dimerize to give a "monomer" of phycocyanin.

[0027] The complete biosynthesis of holo-CpcA was recognized for the first time in the photosynthesizing blue alga *Synechocystis sp.* PCC 6803, and the genes relevant thereto were identified (Nakamura, Y., Kaneko T., Hirosawa M., Miyajima N., Tabata S. (1998) Nucleic Acids Res. 26, 63-67). The heterodimeric PCB lyase is encoded by the genes cpcE and cpcE. Heme oxygenase is encoded by hox1, and ferredoxin oxidoreductase is encoded by the pcyA gene. The cpcA gene encodes the α subunit CpcA and the cpcB gene encodes the β subunit CpcB of phycocyanin.

[0028] Despite the great demand for fluorescent markers having the particular spectral properties of phycocyanin for cell biology and despite the knowledge about the proteins involved in the biosynthesis and the genes thereof, it has not been possible to date to express phycocyanin or functionally active parts thereof - such as, for example, holo-CpcA or holo-CpcB - recombinantly in eukaryotes.

[0029] It is true that heterologous expression of holo-CpcA in *E. coli* has been achieved in the past (Tooley et al. *loc. cit.*). This was done by transfecting the cells with two plasmids having respectively cpcA, cpcE and cpcF gene, and hox1 and pcyA. Induction of gene expression was followed by biosynthesis of holo-CpcA. This holo-CpcA also showed comparable spectral properties like the native holo-CpcA. The yield of holo-CpcA was, however, small because only about one third of the available apo-CpcA was in fact converted into holo-CpcA.

[0030] In their investigations, Tooley *et al.* evaluated whether the cause of the reduced conversion of apo-CpcA into holo-CpcA is possibly a deficiency of available endogenous heme in *E. coli*. This was done by additionally supplying the transformed cells with δ-aminolevulinic acid. However, giving this key compound of heme metabolism in addition did not lead to an increase in the yield of holo-CpcA.

[0031] The authors therefore speculate about whether the codon usage of the genes hox1, pcyA, cpcE and cpcF, which are unfavorable for expression in *E. coli*, might be responsible for the deficient yield. An unequal number of the plasmids respectively present in an *E. coli* cell - which harbor either hox1 and pcyA or else cpcA, cpcE and cpcF - might play a role (concerning this, see also Tooley *et al.*, *loc. cit.*, page 10564, top of right-hand column).

[0032] Other study groups have also long ago recognized the advantages of the use of phycobiliproteins as markers for intracellular processes *in vivo* and therefore attempted to establish systems for providing holo-CpcA in yeasts. For this purpose, for example, the cpcE and cpcF genes for the heterodimeric PCB lyase, and the cpcA gene for coding the apo-α-CPCA were cloned into the yeasts. It was intended, by subsequently giving exogenous PCB, to generate a functional holo CpcA - meaning a functionally active part of phycocyanin - in the yeasts. However, this attempt at semisynthetic provision of holo-CpcA in yeasts failed (Schröder B.G. (1997) Doctorial Dissertation (University of California, Berkeley): Phycobiliprotein: Biosynthesis and Applications).

[0033] The possible reasons for the failed attempt to generate functionally active parts of phycocyanin by adding exogenous PCB to recombinantly produced apo-CpcA in yeasts is discussed in detail by Schröder. Thus, he was able to preclude, through the choice of suitable experimental conditions, that the lack of addition of PCB to CpcA was caused by a possible absence of importation of the external PCB into the cell or else by an active exportation of the initially imported PCB from the cell.

[0034] On the basis of his investigations, he therefore ruled out an inadequate concentration of PCB being the reason for the lack of addition reaction of the CpcE CpcF lyase cloned into the yeast genome. He concluded from his investigations further that an inhibitor which inhibits the addition reaction must be

present in yeast, and for the future refers to the possibility of identifying yeast mutates which do not have this inhibitor.

[0035] Schröder further speculates about the possibility, after elucidation of the complete biosynthetic pathway, of cloning the genes relevant therefor into the yeast and thus achieving a complete *in vivo* biosynthesis (for discussion of this problem, see also Schröder *loc. cit.*, page 77-79). However, Schröder omits to give an unambiguous explanation for the failure of his attempts to provide a semisynthetic biosynthesis.

[0036] It may be pointed out once again in this connection, however, that Tooley *et al.* were, even after elucidation of the biosynthetic pathway, unable to achieve satisfactory expression in *E. coli* (see above).

[0037] With this prior art as background, it is all the more surprising that it has now been possible for the first time to transform a eukaryotic cell which expresses a holo-CpcA as functionally active part of a phycocyanin. In this connection, a "functionally active part" means any part, analog or derivative of phycocyanin which has a structural homology to at least one of the phycocyanin α or β subunits (or parts thereof) and corresponds in its spectral properties substantially to those properties of native phycocyanin.

[0038] In a particularly preferred embodiment, the eukaryotic cell of the invention expresses holo-CpcA. For this purpose, the eukaryotic cell can preferably be transformed with the cpcE and cpcF genes for PCB lyase, with hox1 for heme oxygenase, with pcyA for ferrodoxin reductase and with cpcA for the α subunit. It is thus possible to achieve a completely endogenous biosynthesis of holo-CpcA in a heterologous eukaryotic host.

[0039] However, it is also possible likewise to provide the holo-CpcA by a semisynthetic biosynthetic pathway. It may suffice for this purpose to express the

CpcA or CpcB - or at least the part thereof relevant in each case for addition of the PCB - and the PCB lyase in the eukaryotic cell, and to add PCB exogenously. Alternatively, one or more of the other genes relevant for the biosynthesis may also either be cloned in or else their respective gene product added externally. It is thus possible - depending on the respective specific objective of the experiment - to induce the system in a targeted manner. Subsequent isolation of the holo-CpcA or holo-CpcB from the host cell provides the inventive "functionally active part" of the phycocyanin.

[0040] In the case where both the holo-CpcA and the holo-CpcB are expressed in the cell - either completely endogenously or else semisynthetically as described above - the two subunits may dimerize and form a complete phycocyanin monomer. It is thus possible according to the invention to provide not only a part of the phycocyanin but also a complete phycocyanin.

[0041] The phycocyanin provided according to the invention, or its functionally active parts, have a glycosylation pattern which differentiates them from the respective native protein or the corresponding protein subunit. Thus, for example, the CpcA expressed in a eukaryotic host is glycosylated via an O linkage on the threonine at position 6, the serine at position 10 and the serine at position 162 of the amino acid sequence. These glycosylations are not present in native CpcA.

[0042] A particular advantage of the fluorescent protein of the invention is that its emission spectrum shows a red shift by comparison with its previously known expression forms. At a pH of 8.0, the emission maximum of holo-CpcA is at 643 nm and thus 2 nm further in the longer-wavelength range than, for example, the CpcA expressed in *E. coli*. Although this difference is not quantitatively large, it can serve in practice to distinguish this signal from signals of other fluorescent dyes.

[0043] In a particularly advantageous embodiment of the invention, the holo-CpcA is expressed in a transformed yeast cell. The yeast strain particularly suitable for this is that deposited under admission number DSM 16134 on January 15, 2004, at the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ).

[0044] It has emerged that heterologous expression of holo-CpcA not only depends on successful cloning of the genes necessary for the biosynthesis into the host cell, but can also be influenced by a suitable choice of the conditions for cultivating the host cell. In this connection, cultivating conditions which influence the concentration of available endogenous heme in the host cells may be particularly important. Thus, stimulation of heme biosynthesis or else inhibition of cellular heme degradation may lead to an increase in the endogenous heme concentration.

[0045] The biosynthesis of heme, which comprises a total of 8 steps, has been known for a long time and the genes responsible therefor in yeast have already been identified (e.g. R. Labbe-Bois, P. Labbe: Tetrapyrrol and Heme biosynthesis in the yeast *Saccharomyces cerevisiae*. In: H.A: Dailey (Ed.) Biosynthesis of Heme and Chlorophylls, McGraw-Hill, New York, 1990, 235-286 or A. Chelstowska, J. Rytka: Heme biosynthesis in the yeast *Saccharomyces cerevisiae*. Postby Biochem. 39 (1993) 173-185). The mechanisms regulating this synthetic pathway are not as yet completely understood, however. It has emerged, however, that the heme production rate depends inter alia on the carbon source used in the medium and oxygen availability during cultivation. Thus, addition of glucose leads to a repression of heme synthesis. The same applies to other carbon sources which can be fermented by yeast, such as, for example, galactose. Addition of non-fermentable carbon sources such as ethanol by contrast stimulates heme metabolism 2-3-fold by comparison with glucose (M. Hoffmann, M. Gora, Rythka J: Identification of rate-limiting steps in yeast heme

biosynthesis. Biochemical and Biophysical research Communications 310 (2003) 1247-1253; Chelstowska (1993)).

[0046] It is thus advantageous for the provision according to the invention of the recombinant phycocyanin or its parts to cultivate the transformed yeast cells in a medium with non-fermentable carbon sources in order to increase the endogenous concentration of available heme.

[0047] Alternatively, it is possible in an advantageous embodiment of the preparation method of the invention for inhibitors of heme degradation to be added to the culture medium or cloned into the host organism.

[0048] The enzyme HMX1 is a homolog of heme oxygenase which is lacked by the unmodified yeast (Protchenko O., C.C. Philpott (2003) Regulation of Intracellular Heme Levels by HMX1, a Homologue of heme Oxygenase, in *Saccaromyces cerevisiae. JBC* 278:36582-36587). Since this enzyme is inhibited by iron, it is possible to inhibit this enzyme in a specific culture medium, e.g. by addition of iron.

[0049] The concentration of available heme can likewise according to the invention take place through overexpression of the key enzymes of heme synthesis. Overexpression for this purpose is possible of porphobilinogen synthase, of porphobilinogen deaminase or of uroporphyrinogen III decarboxylase (Hoffmann *et al* (2003) *loc. cit.*).

[0050] It may likewise be advantageous - on use of the ADH promoter as promoter of the genes cloned in - to increase the activity of the promoter, e.g. by raising the glucose concentration. Raising the glucose concentration to increase the promoter activity is admittedly at first sight in conflict with the repression of heme metabolism associated therewith. However, it has been confirmed experimentally that raising the glucose concentration in the medium from the 2%

which is usual for cultivating yeasts for example to 13% in particular leads to an optimized yield of recombinant holo-CpcA with an increase in the fluorescence signal by more than 7-fold (see below).

[0051] In a particularly advantageous embodiment, the culture medium is provided with 2% ethanol. It is known that ethanol on the one hand as non-fermentable carbon source induces heme metabolism (Hoffmann *et al.* (293) *loc. cit.*), but on the other hand also promotes the activity of the ADH promoter (Ruohonen, L., Aalto, M.K., Keränen, S.: Modifications to the ADH1 promoter of Saccharomyces ceriviseae for efficient production of heterologous proteins. Journal of Biotechnology 1995 (39) 193-203) *loc. cit.*).

[0052] It is possible according to the invention to provide, alternatively or supplementarily, for inhibition or delay of the degradation of the recombinant phycocyanin or its parts in the cell. Since a nitrogen deficiency favors increased degradation of the phycocyanin hexamers (Grossman AR (1990) Chromatic adaptation and the events involved in phycobilisome biosynthesis. *Plant, Cell and Environment* 13:651-666), phycocyanin degradation can be counteracted by nitrogen supplementation.

[0053] As already mentioned, in one embodiment of the invention there is transformation of a yeast cell. It is advantageous for this purpose to optimize the PC genes which are cloned in for expression in yeast by first adjusting the GC content of the genes to an average value of 45%. It is thus possible to prolong the genetic stability and the half life of the mRNA in the host organism.

[0054] The efficiency of expression can additionally be increased further by adjusting the PC genes to the particular host organism in relation to their codon usage. In a particularly preferred embodiment, the eukaryotic cell is transformed with one or more plasmids which one or more of the oligonucleotide sequences of SEQ ID No. 6-10.

[0055] For successful transformation of the eukaryotic host, the selected PC genes are initially cloned into a cloning vector and can then be stably integrated via an integration vector into the genome of the host cell. The genes may in this case be distributed over one or more cloning vectors. In a particularly preferred embodiment of the invention, all 5 PC genes are cloned in only one plasmid.

[0056] If constitutive expression of a gene on the vector or in the genome is desired, a constitutive promoter - for example an ADH promoter - can be put upstream of this gene. However, inducible promoters such as, for example, the GAL promoter are also possible. The promoters may either be put upstream of each individual gene or together regulate the complete group of genes.

[0057] In a preferred embodiment, an ADH promoter was chosen in order to enable expression of the PC genes hox1, cpcE, pcyA and cpcF. The ADH promoter is in this case put upstream of each gene in each case. In this embodiment, the cpcA gene is regulated by the GAL promoter. It is thus possible to create an inducible system for the biosynthesis of the phycocyanin or of the holo-α-phycocyanin.

[0058] It is, of course, also possible alternatively to place the cpcA gene under the control of the ADH promoter and one or more of the other genes under the control of the GAL promoter or of any other promoter.

[0059] If an ADH1 promoter is used, it is advantageous to use the ADH promoter not in its original length of about 1500 bp, because experience shows that the activity of the ADH1 promoter declines over time. It was thus possible to observe that the complete promoter switches off in the late exponential growth of the cells.

[0060] According to investigations by Ruohonen *et al.*, this property of the ADH1 promoter is attributable to a 700 bp fragment located upstream. Deletion of

this region as far as an "intermediate promoter" of about 700 bp is by contrast suitable for high and stable expression of the cloned genes. The publication by Ruohonen *et al.* is incorporated by reference in its entirety relating to the structure and location of an ADH1 promoter which can be used according to the invention.

[0061] In a preferred embodiment, an ADH1 promoter with the nucleotide sequence shown in SEQ. ID. No. 5 is employed.

[0062] Since eukaryotic mRNA is, in contrast to prokaryotic mRNA, not polycistronic, the genes which have been cloned in must in each case be provided with a termination sequence. In an advantageous embodiment, the termination sequence of Guo and Sherman is employed (Guo, Z., Sherman, F. Signals sufficient for 3' end formation of Yeast mRNA. Mol. Cell Biol 1996 (16), 2772-2776). The termination sequence of SEQ. ID. No. 13 which is depicted under number 3 is particularly preferred.

Exemplary embodiment

1. Cloning strategy

(see also Fig. 1 "Cloning strategy: construction and cloning of the genes/vectors")

1.1. Cloning and synthesis of the genes:

Since the neighborhood of the start codon is to be retained, the following enzymes are suitable:

Enzyme	Recognition seq.	Isoschizomers	Compatible ends	Recut
BspHI	T|CATG A A GTAC|T		NcoI AflIII (A|CATGT)	NlaIII
NcoI	C|CATG G G GTAC|C		AflIII BspHI, PciI	NlaIII

BspHI and NcoI are compatible and do not cut in the gene sequences.

[0063] The genes were cloned into the vectors via 5' NcoI or BspHI and 3' XhoI. For this purpose, the genes received 5' the restriction cleavage site appropriate in each case according to the neighborhood of the start codon. The restriction cleavage site for XhoI was attached at the 3' end. Care was taken here that the minimum number of amino acids were additionally inserted into the sequence through the cloning. In addition, the attachment of the XhoI cleavage site was not to result in a frame shift. Account was taken of this by inserting the bases AA in front of the cleavage site and appending an A after the cleavage site. The amino acid sequence NSR was thereby appended to each gene, ensuring that it is not encoded by rare codons.

1.2. Cloning and synthesis of the vectors:

The vectors are intended to permit subsequent excision of the complete gene optionally with and without His tag in order to enable expression of the individual genes to be detected. However, the His tag was omitted for cloning the genes consecutively, in order to avoid possible problems with the additional sequence.

[0064] Thus, the vectors comprise NcoI-XhoI or BspHI-XhoI in each case with and without His tag and with and without termination sequence. The present design of the vectors makes it possible for the genes to be cloned therein and then excised with/without His tag, stop codon and termination sequence, and cloned further.

[0065] The initial vector used, into which the synthesized MCSs and the genes were cloned consecutively, was pSL 1190 (see map of pSL 1180 (Fig. 2); the vectors pSL 1190 and pSL 1180 differ only in the orientation of the MCSs). The restriction cleavage sites which were taken into account in the synthesis of the four different MCS sequences are identified by boxes in the figure - more possibilities are thus present for the cloning. The restriction cleavage sites

eventually used in the cloning of the PC genes are marked in the figure additionally by underlining of the respective boxes.

[0066] The following things were considered in the design of the MCSs for the vectors:

- the enzymes used must not cut in the promoter sequences
- they must not cut in the backbone of the vector
- the distance between the restriction cleavage sites must be large enough for adjacent enzymes both to be able to cut
- it must be possible to clone the MCSs into pSL 1190 - the EcoRI and HindIII enzyme cleavage sites are provided at the ends of the MCSs for this purpose
- cleavage sites must be available for integrating the complete PC cassette (all 5 genes cloned consecutively) into the vector pRS306 (here: SacI/SalI)
- the possibility of linearization in the URA3 gene (selection marker) must exist in the vector pRS306 so that the vector can integrate and the cassette is genomically integrated.

The sequences of the MCSs of the vectors are appended (SEQ. ID. No. 1 - SEQ. ID. No. 4).

[0067] It is necessary, in order to be able to carry out a cloning with BspHI, to delete from the vectors 656 and 660 (pSL 1190 with synthesized MCS which comprises a BspHI site) three cleavage sites of this enzyme from the backbone of the vector. The Quik Change Multisite directed mutagenesis kit from Stratagene is employed for this purpose, and the success of the experiment is verified by restriction analyses of the resulting vectors.

Clones resulting from the vector cloning: (the clones used further are indicated in **bold**)

Clone No.: B	Clone description
647	GeneArt mcs-A = 03-200; NcoI/His tag in PCR-Script; **647**
648	GeneArt mcs-B = 03-201; BspHI/His tag in PCR-Script; **648**
649	GeneArt mcs-C = 03-215; NcoI/without His tag in PCR-Script; **649**
650	GeneArt mcs-D = 03-216; BspHI/without His tag in PCR-Script; **650**
651 to 653	GeneArt mcs = 03-199 in pRS306; **651**
654, 655	GeneArt mcs-A = 03-200 = NcoI/His tag in pSL 1190, **654**
656, 657	GeneArt mcs-B = 03-201 = BspHI/His tag in pSL 1190 **656**
658, 659	GeneArt mcs-C = 03-215 = NcoI/without His tag in pSL 1190; **658**
660, 661	GeneArt mcs-D = 03-216 = BspHI/without His tag in pSL 1190; **660**
682 to 689	QuikChange with B656 for deleting BspHI sites; **684, 685**
690 to 697	QuikChange with B660 for deleting BspHI sites; **690, 693**

1.3. Choice and cloning of the ADH promoter:

The ADH promoter (alcohol dehydrogenase) is present in many cloning vectors and there leads to constitutive expression of downstream genes. However, on use thereof, there is often found to be a decline in the expression rate of the cloned gene until expression is completely switched off after about 3 days.

[0068] However, it has emerged that an ADH promoter comprising a 700 bp-long piece of the original (complete) promoter leads to expression of the cloned genes throughout the growth cycle of the yeast (Ruohonen et al. (1995) *loc. cit.*).

[0069] This piece of the promoter was the starting point for the ADH promoter used according to the invention, although the primers chosen for the PCR in the cloning of the promoter were such that an ADH fragment totaling 719 bp resulted (see SEQ. ID. No. 5).

[0070] The synthesis of the ADH promoter was carried out by means of PCR from the genome of *S. cerevisiae* Y190. The cloning took place in a three-fragment ligation with the gene in question and the vector together. The same reverse primer having an NcoI cleavage site was always employed for this purpose, because NcoI and BspHI are compatible. The forward primer has in each case another appended restriction cleavage site which is selected according to the cloning step.

[0071] Sequences of the primer used (the restriction cleavage site is underlined in each case): the additional bases are required for the enzyme to be able to cut the PCR fragment:

ADH-SacI-for	CAATTAGAGCTCATATCCTTTTGTTGTTTCCGGGTG
ADH-SacII-for	ATCCCCGCGGATATCCTTTTGTTGTTTCCGGGTG
ADH-NotI-for	AAGGAAAAAAGCGGCCGCATATCCTTTTGTTGTTTCCG GGTG
ADH-XbaI-for	AGCTCTAGAATATCCTTTTGTTGTTTCCGGGTG
ADH-BglII-for	CGAAGATCTATATCCTTTTGTTGTTTCCGGGTG
ADH-NcoI-rev	GTAAGCCATGGTGTATATGAGATAGTTGATTGTATGCTTGG

2. Optimization of codon usage

The codon usage was optimized for the frequency of codon usage of *S. cerevisiae*. The GC content was where possible adjusted to an average value of about 45% in order to increase the genetic stability of the mRNA half life.

In addition, the following restriction cleavage sites intended for the cloning were deleted from the genes:

Gene	Restriction enzyme	Position (bp)
CpcA	KpnI	312
CcpE	NheI	779
	YbaI	493
	HpaI	366
	HpaI	549

The following sequences were used for cloning of the genes required for expression of the holo-CpcA ("PC genes"):

Gene	SEQ. ID. No.
cpcA	SEQ. ID. No. 6
cpcE	SEQ. ID. No. 7
cpcF	SEQ. ID. No. 8
hox1	SEQ. ID. No. 9
pPcyA	SEQ. ID. No. 10
cpcB	SEQ. ID. No. 11

If it is intended to express a holo-CpcB it is possible to clone in the cpcB gene (SEQ. ID. No. 11) as alternative to cpcA. The gene encodes an amino acid sequence shown in SEQ. ID. No. 12.

3. Choice of the termination sequence

Since eukaryotic mRNAs are not, in contrast to prokaryotic ones, polycistronic, a termination must take place after each gene. Guo and Sherman (1996, *loc. cit.*) describe a synthetic DNA sequence which is sufficient to form a 3' end and consists of efficiency element, positioning element and poly(A) region.

[0072] For our purposes, the XbaI cleavage site was deleted, and the spacer region between efficiency element and positioning element was chosen as follows: ACTCTGTAGA (instead of ACTGTCTAGA; XbaI cleavage site (SEQ. ID. No. 13)).

4. Cloning of the individual genes and test of the expression of the individual genes

The procedure for the cloning of the individual genes and for detecting their successful expression is depicted diagrammatically in Fig. 3. The depiction relates by way of example to the clone B654:pSL 1190 with NcoI cleavage site and His tag.

[0073] The finished vector was linearized with StuI. Transformation took place into the yeast strain Y190 with URA deficiency.

[0074] Firstly, preliminary tests were carried out with two other fluorophores in order to test whether expression of the genes takes place behind the chosen ADH promoter and with the termination sequence employed.

[0075] For this detection, the genes were each cloned singly with the ADH promoter into the vector pRS306. This is an integration vector, so that the genes can subsequently be integrated into the genome of the yeast. The situation thus corresponds substantially to the intended aim, namely for the five genes relevant for the holo-CpcA ("PC genes") to be present consecutively. SacI/SalI was chosen as cleavage sites for integrating the gene with promoter into the vector pRS306, because foreign sequence can be introduced the most into the yeast in this way.

[0076] For this purpose, the genes were first cloned into the appropriate vector (pSL 1190 with BspHI or NcoI site and His tag) in order subsequently to be ligated in a three-fragment ligation with the ADH promoter.

[0077] To carry out the three-fragment ligation, the amounts of DNA employed were determined using an agarose gel with included marker. The amounts of DNA employed were then calculated at the molar level. 25 fmol were employed to the vector, and 125 fmol to each of the two inserts.

[0078] The cell pellet after the cultivation of protein-expressing yeasts was worked up with glass beads and vortexer. Denaturing conditions were chosen in this case. The instructions of the manufacturer of the Ni-NTA columns were complied with in the purification of the His-tagged protein.

5. Cloning of the genes for holo-CpcA consecutively

The procedure for cloning the PC genes consecutively is depicted diagrammatically in Figs. 4a to 4c. The map of the vector with the PC genes cloned in is shown in Fig. 5. SEQ. ID. No. 14 comprises the sequence data for the vector. Where the Gal promoter is upstream of the CpcA, the vector shows a structure in accordance with Fig. 6.

[0079] The genes were cloned in pSL 1190 with synthesized MCS without His tag. These genes were then excised with appropriate enzymes and put together with the ADH promoter consecutively into the vector pSL 1190. The success of the clonings was in this case detected by means of bacterial colony PCR for the respective gene, and the identity of the isolated DNA was verified by test cuts. In addition, the promoter portion of the vectors resulting in each case was sequenced (AGOWA) because this was generated by PCR (varying forward primer with different restriction enzymes) and errors might be present. The proofreading polymerase Pfu (Promega) was employed for these PCR applications.

[0080] Once the genes were present consecutively in the vector pSL 1190, the complete cassette was cloned via the SacI/SalI restriction cleavage sites into the integration vector pRS306. In this cloning step, the cpcA gene was detected with appropriate primers, and the vector pRS306 was detected by means of a PCR with primers which amplify a piece from the URA3 selection marker, in the transformed bacterial colonies.

Sequence of the primers used for detecting the cloned genes/integration vector pRS306:

		Length of the fragment
cpcF-for	TGATTCAAGCTGTTGAAACC	361 bp
cpcF-rev	GATTGCCATCTCAAGAAACC	
hox-for	CTGCTGGTCAAGCATACG	413 bp
hox-rev	GGTAGCCAAACCAACTTCG	
cpcA-for	CTTTCGGTAGATTGAGACAAGC	414 bp
cpcA-rec	CAAAGCGTTAATAGCGTAATCC	
pcyA-rev	TTGACCAACTCTTCTTTGATGC	579 bp
pcyA-rev	AGCTTCAGACAATGGTTCAGC	
cpcE-for	AAGATGAAACCGATAGATCACC	609 bp
cpcE-rev	AAGCCTTAGATTCAGAAGATGC	
URA-pRS-for	GAACGTGCTGCTACTCATCC	613 bp
URA-pRS-rev	CGTCTCCCTTGTCATCTAAACC	

[0081] In order to be able to detect the holo-protein by means of Western blotting and immunodetection for the His tag, the cpcA gene with His tag was placed behind the cassette with the remaining genes, and then this complete cassette was cloned into pRS306. After transformation with subsequent expression of the protein in yeast, the protein was purified and visualized in a Western blot (Fig. 7).

Western blot:

[0082] The primary antibody employed for the Western blot was the *mouse antihistidine tag* (Serotec) 1:250. The secondary antibody used was *goat antimouse alkaline phosphatase* (Dianova) 1:5000. Detection took place with BCIP/NBT blue liquid (Sigma).

The prestained SDS gel marker (Invitrogen) was used as marker.

6. Spectrum of the holo-CpcA of the invention

The spectrum (Fig. 8) of the holo-CpcA of the invention was recorded in the following buffer:

Lysis buffer:

50 mM NaH_2PO_4

300 mM NaCl

10 mM imidazole

pH 8.0

Elution buffer:

50 mM NaH_2PO_4

300 mM NaCl

250 mM imidazole

pH 8.0

The lysis buffer and the elution buffer were present in the ratio 60% to 40%.

7. Yeast strain and cultivation conditions

A uracil-deficient variant of the Y190 yeast strain was used as host organism for the heterologous expression according to the invention of the PC genes. The yeast strain Y 190 showed originally the following genotype: MATa, leu2-3, leu2-112, ura3-52, trp1-901, his3-Δ200, ade2-101, GAL4-Δgal80-Δ, URA3::GAL-LacZ, LYS::GAL-HIS3, cyh^r.

[0083] To provide a uracil-deficient Y190 yeast, yeast cells which show a deletion of the URA3 gene were selected from this strain by the method of Boeke *et al.* [Boeke, J., Lacroute, F. and Fink, G.R (1984). A positive selection for mutants lacking orotidine-5-phosphate decarboxylase activity in yeast: 5-Fluoro-orotic acid resistance. Mol. Gen. Genet 197, 345-346].

[0084] These URA3 deletion mutants were cultivated on YPAD medium or on any synthetic complete media. It has proved to be particularly advantageous to use a glucose concentration of up to 13% instead of the 2% glucose concentration proposed in the literature. However, all the concentrations between 2 and 20% are possible.